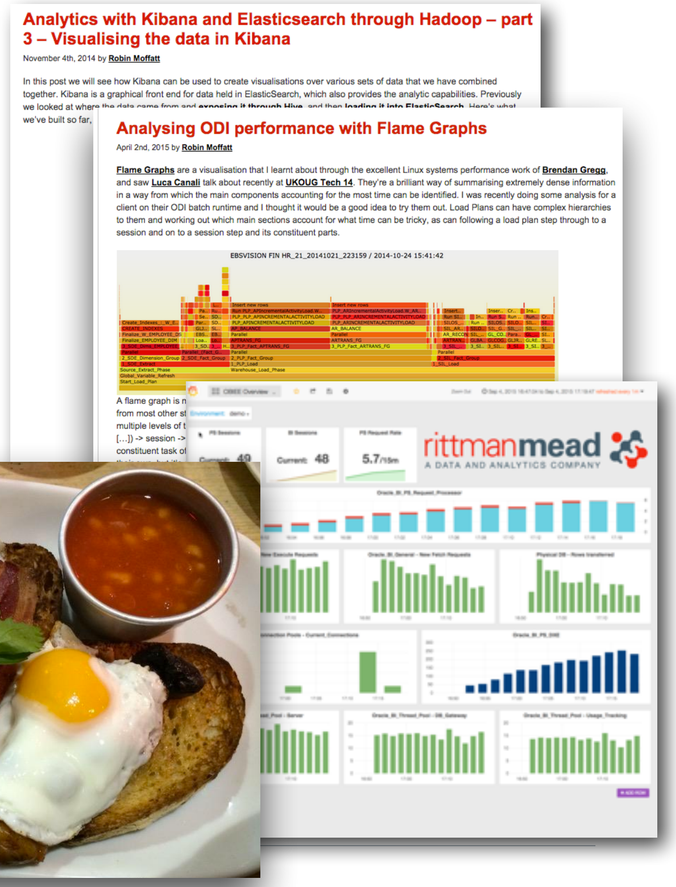
Data Discovery & Systems Diagnostics with the ELK Stack

Robin Moffatt, Principal Consultant, Rittman Mead | YoDB November 2015

Robin Moffatt

- Principal Consultant with Rittman Mead
  - OBIEE & ODI / SysAdmin / Performance
- Previously …
  - OBIEE/DW developer at large UK retailer
  - SQL Server DBA, BusinessObjects, DB2, COBOL….

ORACLE ACE



- Blog: **http://ritt.md/rmoff**
- Twitter: **@rmoff**
- IRC: **rmoff** / #obihackers / freenode

ELK

ELK

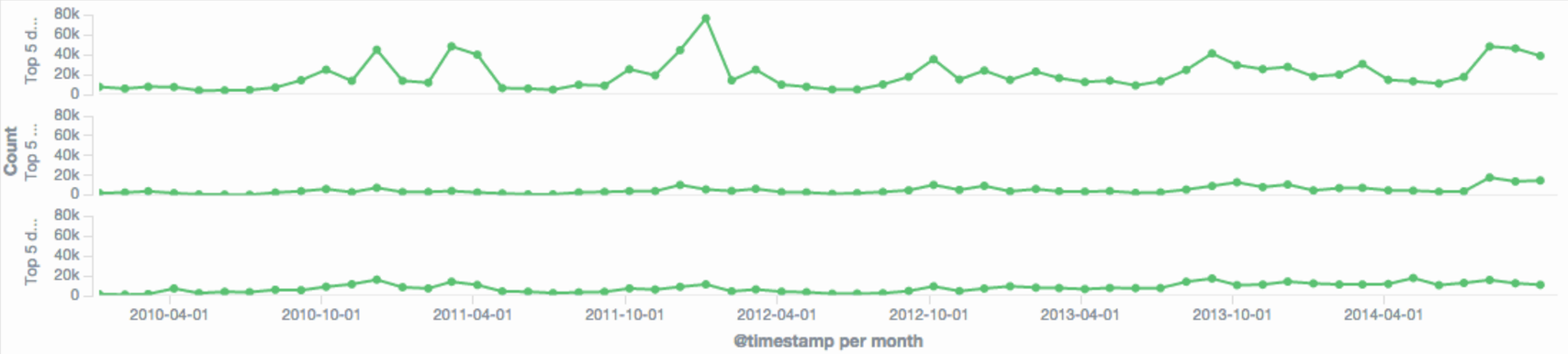
- **E**lasticsearch - schema-free, distributed data store
- **L**ogstash - centralised data processing
- **K**ibana - analytics and visualisation

elastic

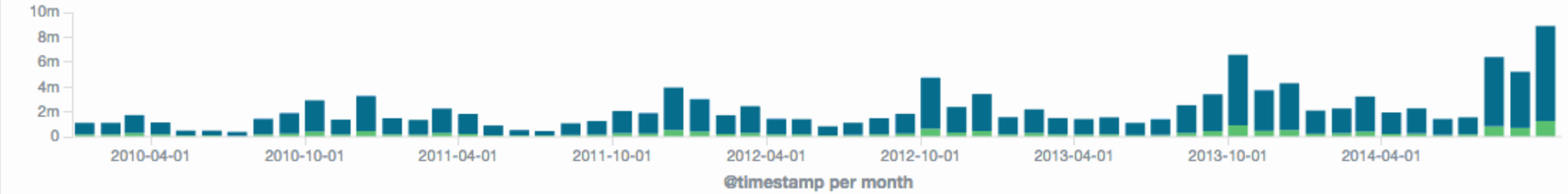
Data Discovery

ELK - Data Discovery

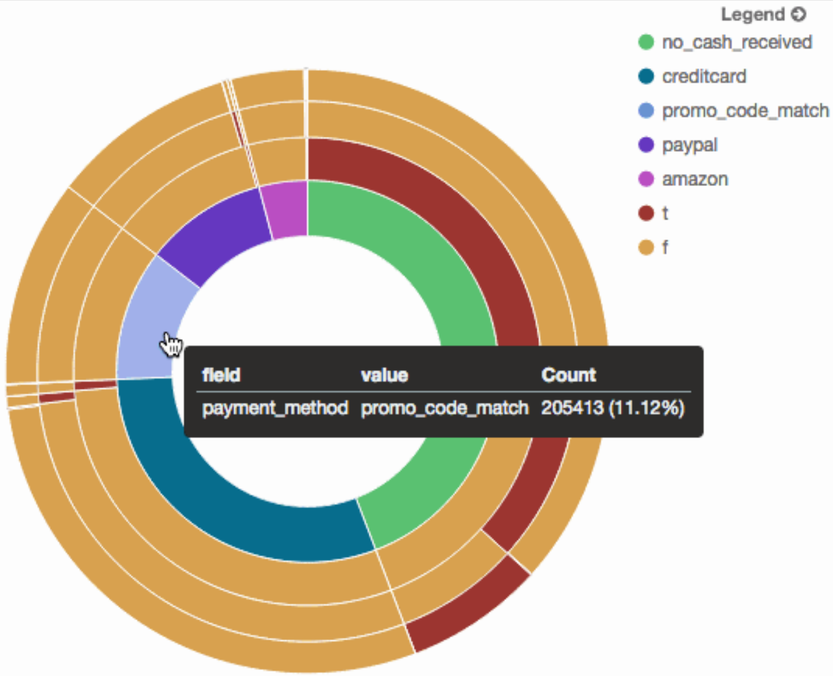
Donation count by value range



Amount given over time.



Payment Methods



Donations

« 1 2 3 4 5 ...10 »

Time	donation_message	donation_total	donor_city	donor_state	payment_method	payment_was_promo_matched	via_giving_page
October 31st 2014, 22:55:01.662	"Claire Olivia Ryan has the best teacher and a wonderful homeroom mother....Lisa Michele Ryan. "	50		CA	creditcard	f	f
October 31st 2014, 22:54:16.076	"I love that Ms.D wants the best for her students! I thankful I had wonderful teachers like her!"	200		TX	paypal	f	t
October 31st 2014, 22:53:04.329	"J.Crew is proud to match Associate donations to support this classroom."	197.19		NY	promo_code_match	f	t
October 31st 2014, 22:52:22.228	"I want the students to get the most out of science."	5			paypal	t	t
October 31st 2014, 22:51:04.836		50		OR	creditcard	f	f

Systems Diagnostics

ELK - Systems Diagnostics

Session Overview

*

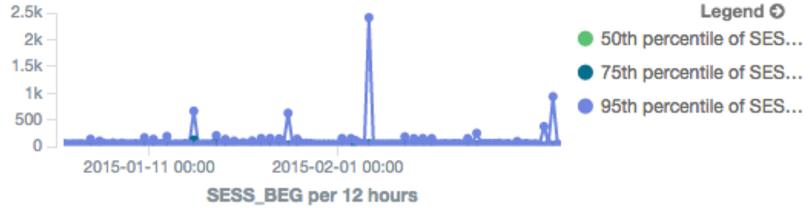
_type: "ec" Actions

Source

Top 20 _type	Count
ec	127544

Session Count

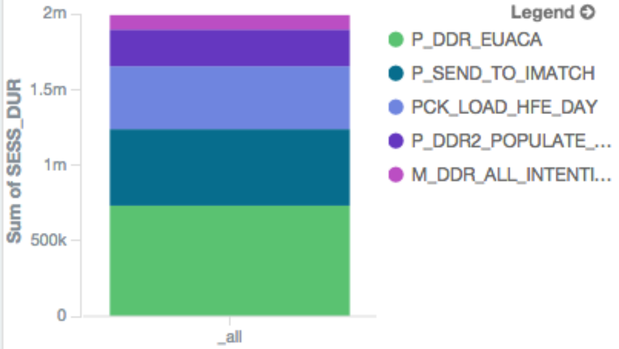
Session duration



Heading

Rittman Mead
ODI Monitoring

Top five sessions (cumulative runtime)



Top five sessions (cumulative duration)

Top 5 SESS_NAME	Count	Sum of SESS_DUR	50th percentile of SESS_DUR	95th percentile of SESS_DUR
P_DDR_EUACA	12353	731735	51	86.832
P_SEND_TO_IMATCH	11367	505404	36	58.878
PCK_LOAD_HFE_DAY	145	415981	1768	9663.7
P_DDR2_POPULATE_HISTORY	82	242826	1025	9522.25
M_DDR_ALL_INTENTIONS_HIST	82	95244	673	976.65

Export: Raw Formatted

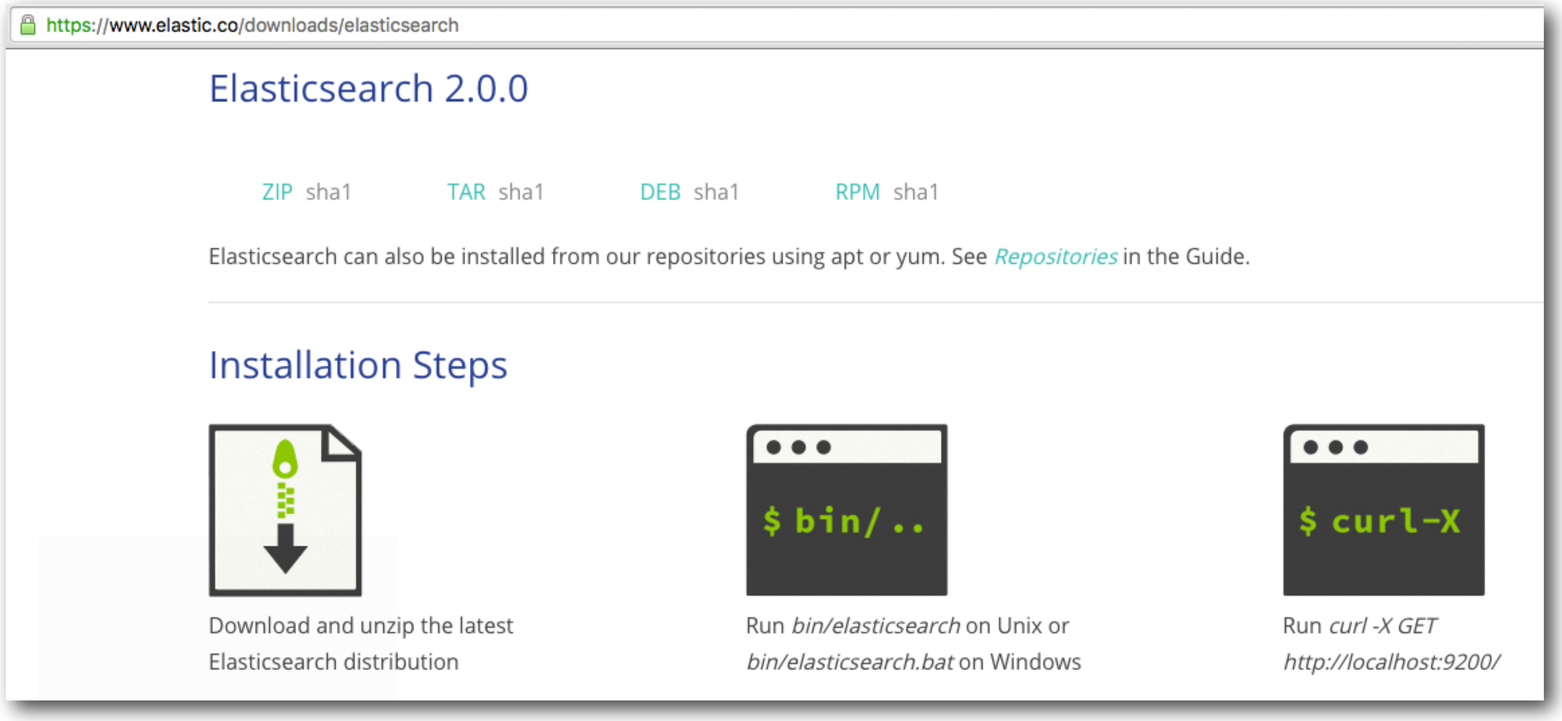
Top sessions by mean runtime

Top 50 SESS_NAME	50th percentile of SESS_DUR	95th percentile of SESS_DUR	Sum of SESS_DUR	Count
M_DDR_ALL_INTENT_0513_POP	7552	9232.95	61669	8
P_POPULATE_CNS_DEP_DATA	5069	5069	10138	2
M_DDR_SCHEDULE_030_LD_IMATCH	2171.5	3014	8686	4
PCK_LOAD_HFE_DAY	1768	9663.7	415981	145

Top five sessions (cumulative num rows)

Getting started is easy!

1. **Download**
2. **Unarchive**
3. **Run**
4. *…*
5. *erm*
6. *that's it!*



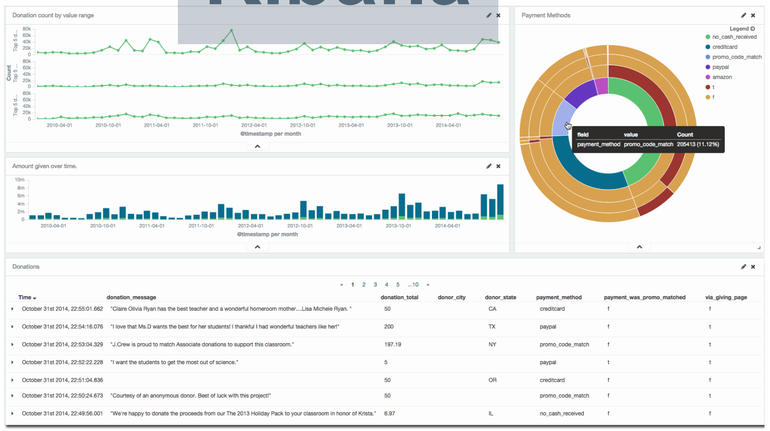
```
78.4.185.138 - - [05/Apr/2015:09:28:44 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http://
159.20.180.99 - - [05/Apr/2015:09:28:44 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http:/
78.5.211.46 - - [05/Apr/2015:09:28:45 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http://w
78.6.58.90 - - [05/Apr/2015:09:28:45 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http://ww
93.51.248.26 - - [05/Apr/2015:09:28:45 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http://
78.7.30.14 - - [05/Apr/2015:09:28:45 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http://ww
78.6.6.6 - - [05/Apr/2015:09:28:46 +0000] "POST /wp-login.php HTTP/1.1" 500 200 "http://www.
```

logs

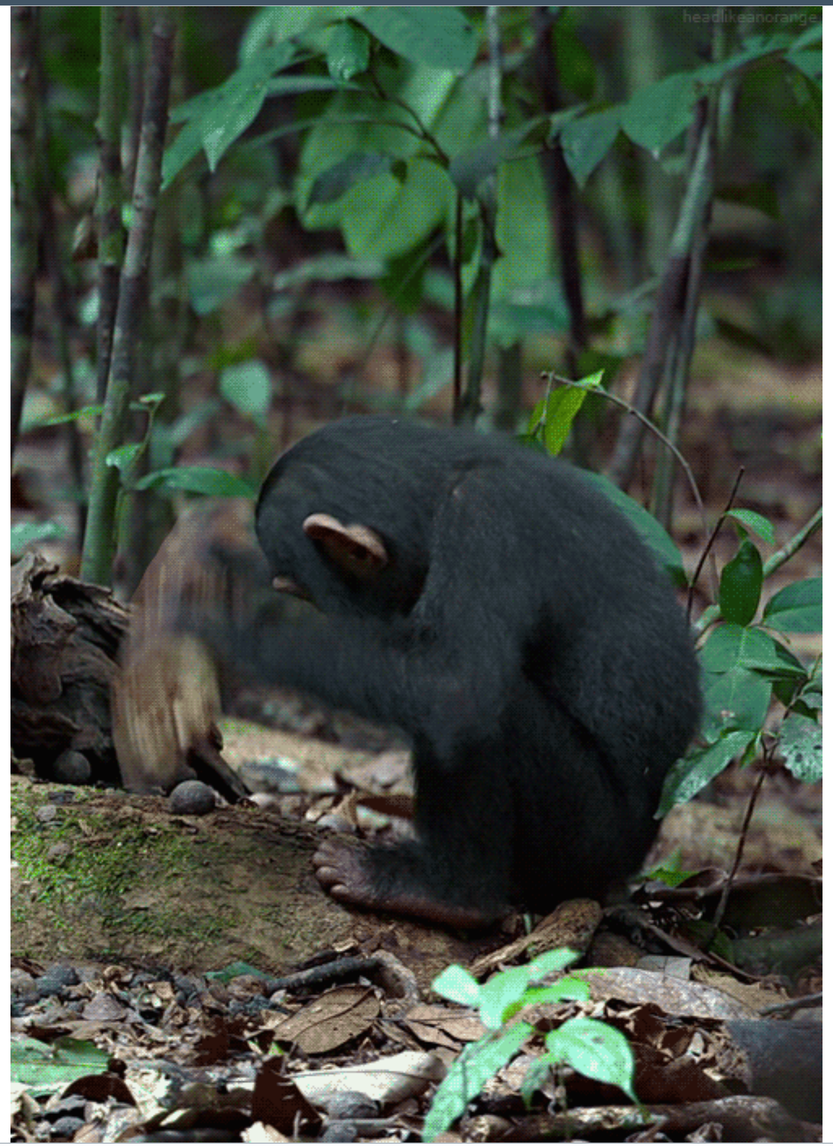
logstash

Elasticsearch

Kibana



DEMO!



twitter

```
"created_at" => "Thu May 07 10:34:45 +0000 2015",
"id" => 596261881842671616,
"id_str" => "596261881842671616",
"text" => "I need one more volunteer to argue agai                    being the answer for users' pra
"source" => "<a href=\"http://twitter.com/downloa                    ollow\">Twitter for iPhone</a>",
"truncated" => false,
"in_reply_to_status_id" => nil,
"in_reply_to_status_id_str" => nil,
"in_reply_to_user_id" => nil,
"in_reply_to_user_id_str" => nil,
"in_reply_to_screen_name" => nil,
```

Logstash

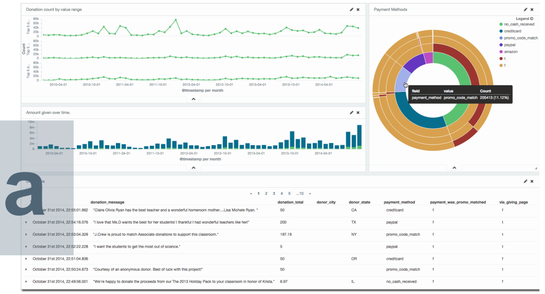
```
filter {
    if [text] =~ /(freelance|job|training)/ {
    drop{}
    }
}
```

```
output {
        if [message] =~ /Labour/ {
                csv { fields => ["@timestamp"
        }
        elasticsearch { host => "localhost" i
        stdout { codec => rubydebug }
```

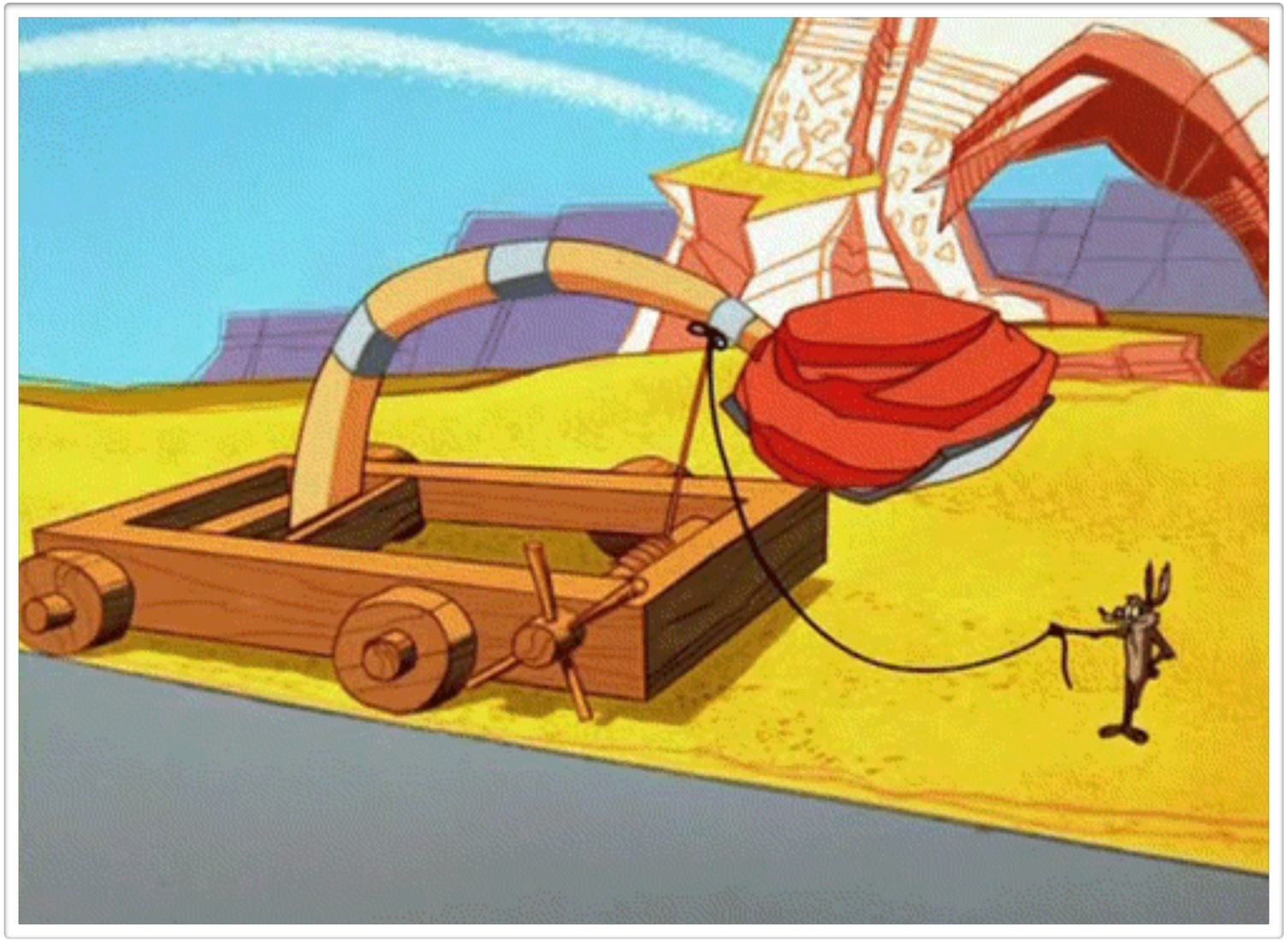
csv

Elasticsearch

Kibana



DEMO!



Elasticsearch

- **The core component of the ELK stack**
- Based on Apache Lucene (same as Cloudera's Solr)
- Distributed for scalability & resilience
- Near-realtime document indexing

elasticsearch

elastic

Elasticsearch Uses - Search and Analytics

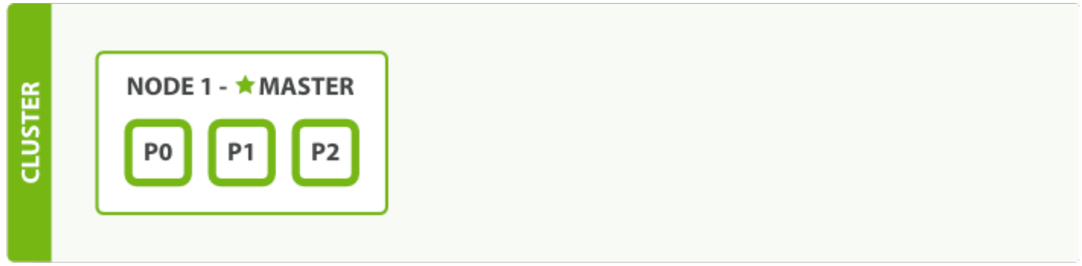
- **Search**
  - Soundcloud
  - GitHub

- **Analytics**
  - The Guardian's Ophan application

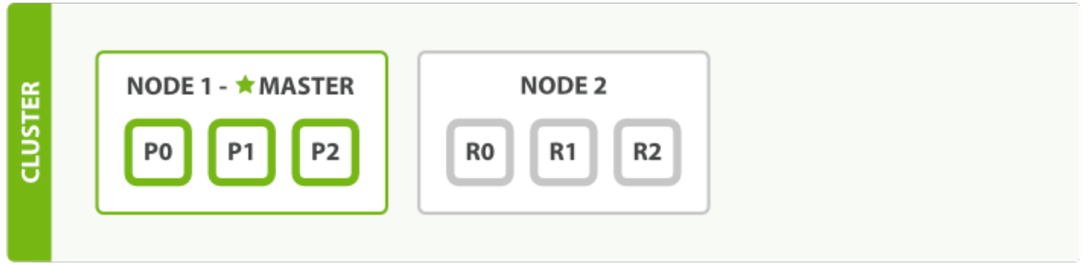
http://tnw.to/s3NV5
https://www.elastic.co/assets/bltd061cc55096a5780/case-study-the-guardian.pdf

Elasticsearch

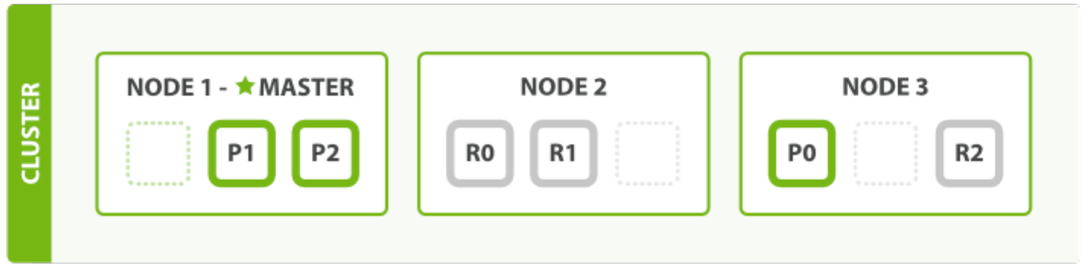
- Stores data as JSON **documents** within an **index**



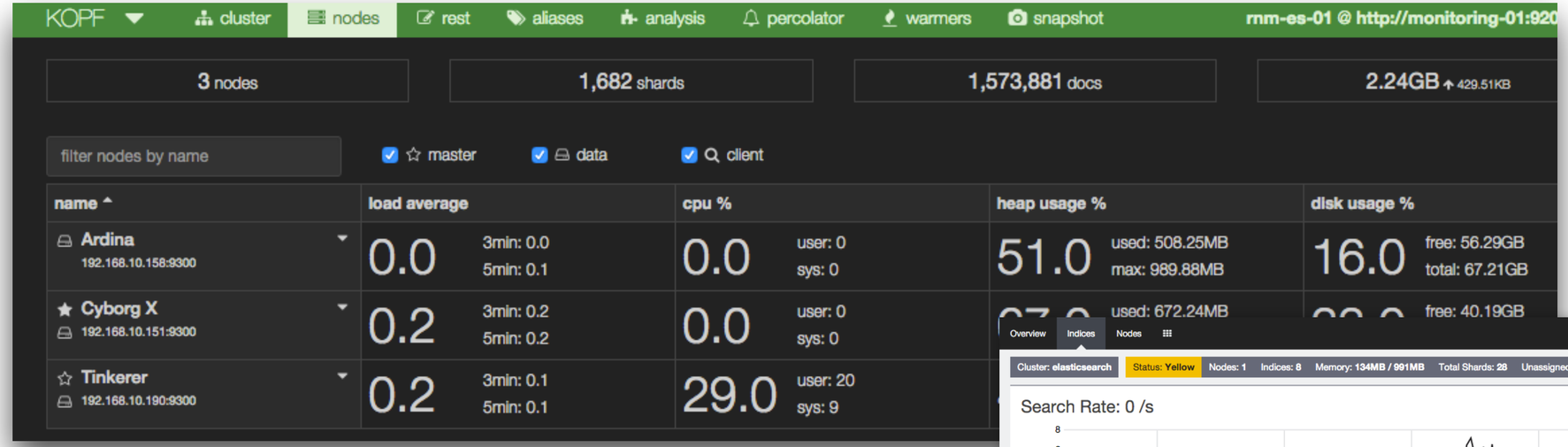
- An index is made up of **shards**



- Shards are distributed around a cluster **automatically**
  - Resilience and scale-out are simple

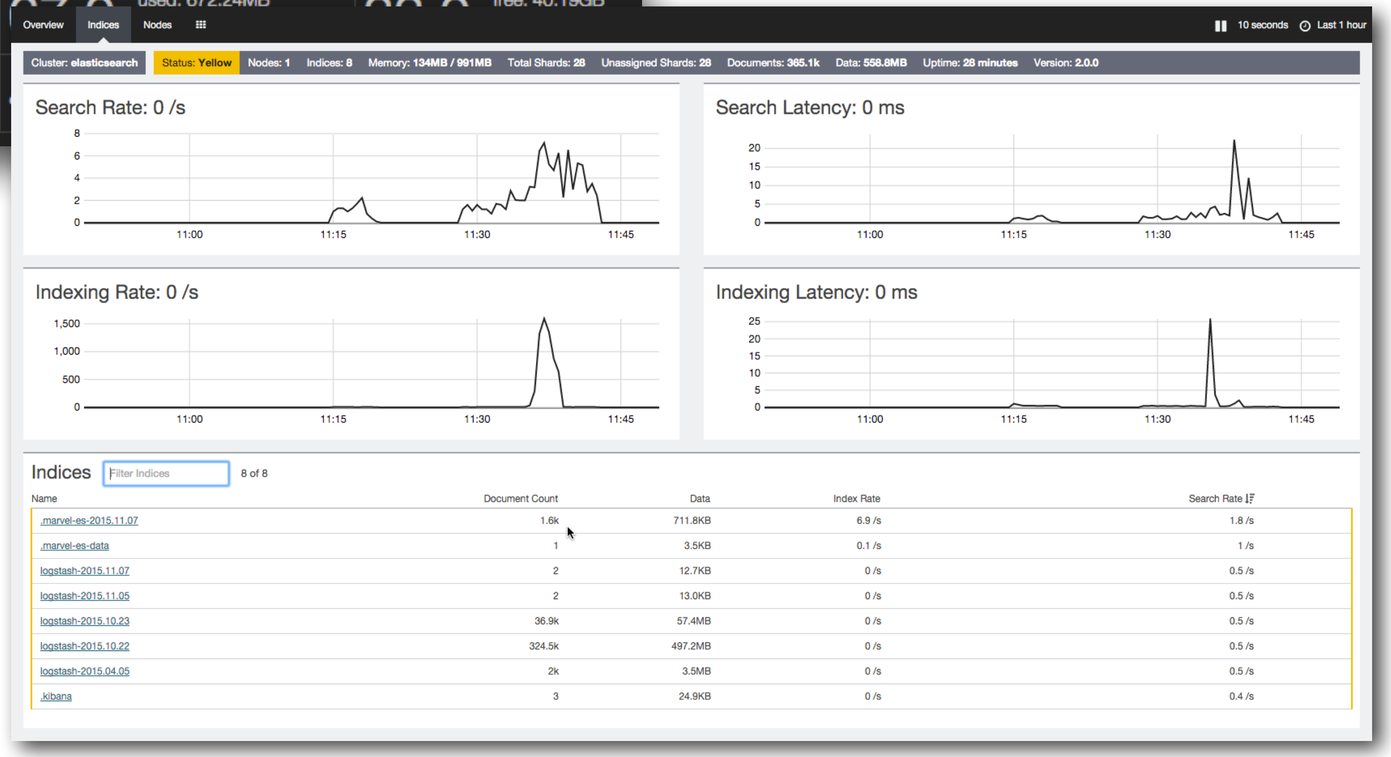


Elasticsearch Administration



Kopf

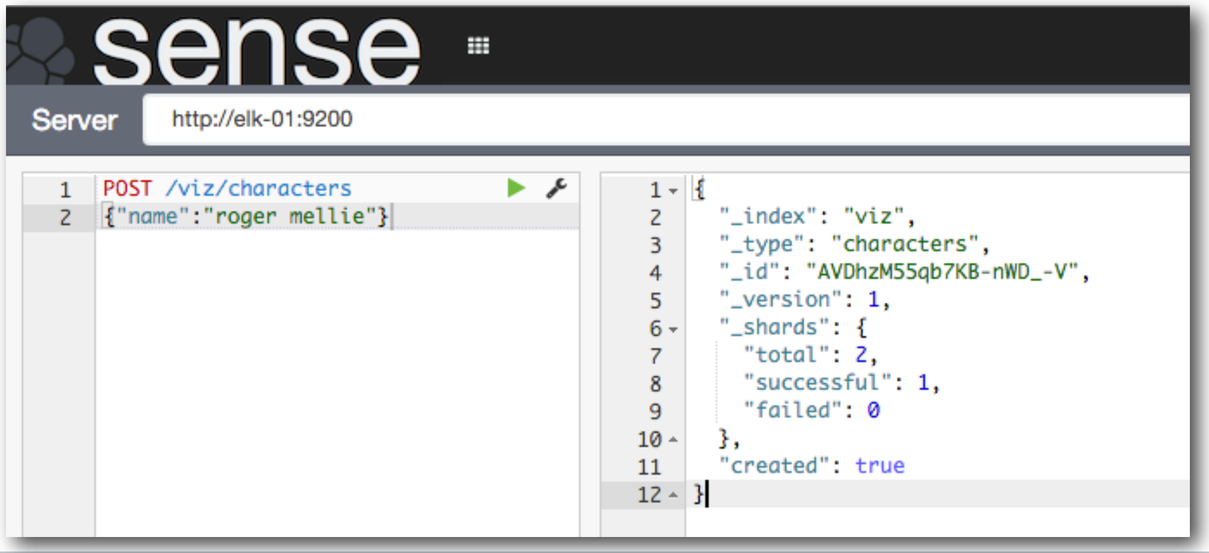
https://github.com/lmenezes/elasticsearch-kopf



Marvel

Working with the REST API

- **curl** is all you need ;-)
  - optionally with **jq** for syntax highlighting of the resulting json
- **Sense** is a free plugin from Elastic for Kibana and useful for rapid prototyping of more complex interactions



Elasticsearch REST API

```
$ curl -XPOST
'http://es:9200/viz/characters/'
-d '{"name":"finbarr saunders"}'
```

Elasticsearch REST API

```
$ curl -XPOST
'http://es:9200/viz/characters/'
-d '{"name":"roger mellie”,
 "notes":"the man on the tele"}'
```

Elasticsearch REST API

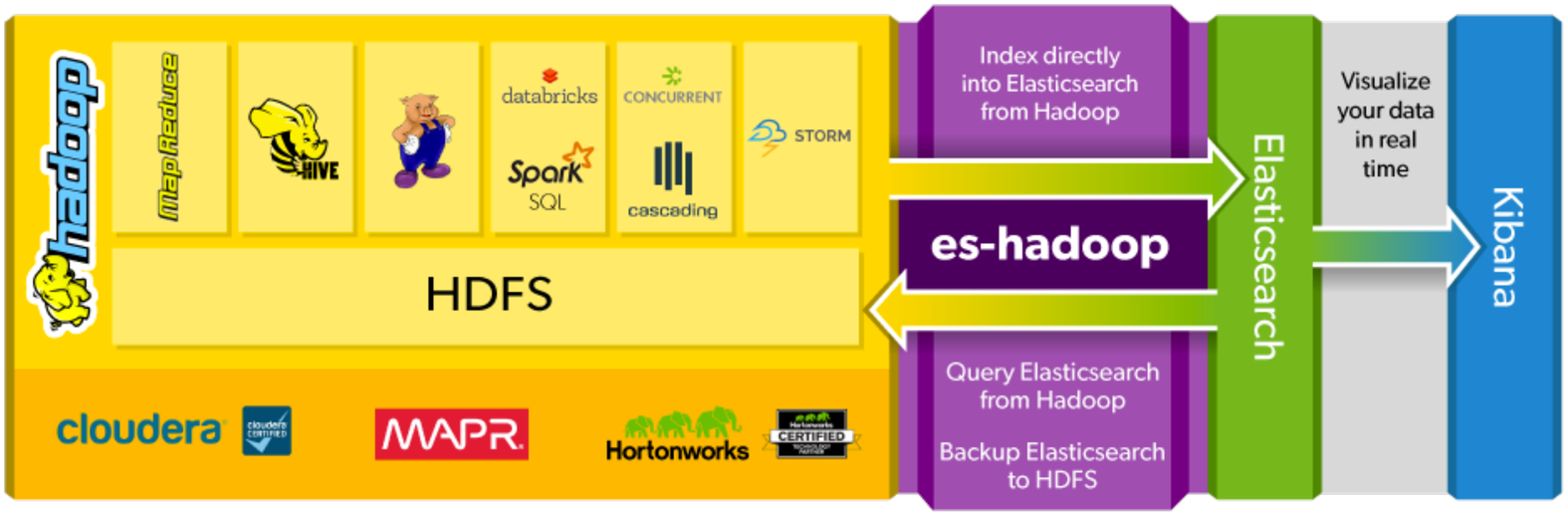
```
$ curl -XGET 'http://localhost:9200/viz/_search?q=roger'
[…]
  "hits" : {   "total" : 1,    "max_score" : 0.11506981,
    "hits" : [ {
      "_index" : "viz",
      "_type" : "characters",
      "_id" : "AUyyNUrTI0Rm5Pb-t8_l",
      "_score" : 0.11506981,
      "_source":{"name":"roger mellie" ,"notes":"the man on
                                                the tele"}
```

Elasticsearch REST API

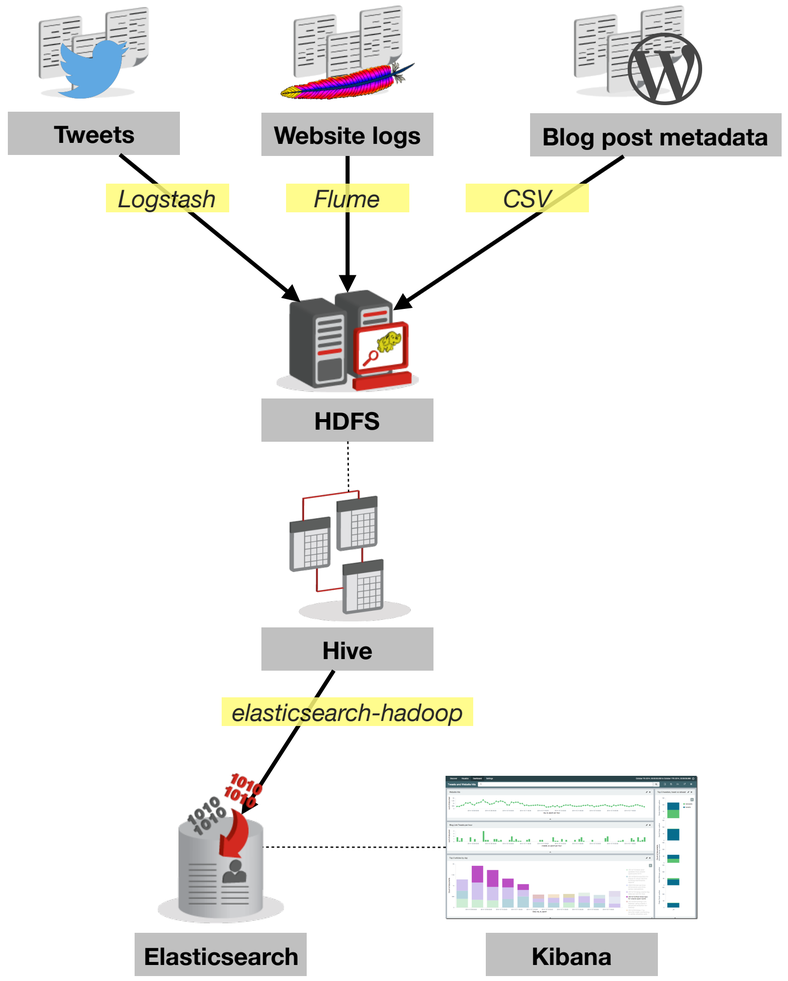
```
$ curl -XDELETE
      'http://localhost:9200/viz'
```

Elasticsearch-Hadoop

- Two-way connector between Hadoop and Elasticsearch
- Read/Write with Elasticsearch from Hive, Pig, Spark, etc



https://www.elastic.co/products/hadoop



http://ritt.md/elk-hadoop-01

Logstash

input
log
csv
json
kafka
stdin
syslog
tcp
jdbc
tsv
twitter
filter
grok
geoip
mutate
drop
output
elasticsearch
email
nagios
kafka
pagerduty
stdout
file

Logstash

- ***Does Logstash support <foo>*** …. yes, probably!
  - Vast number of supported input (and output) formats

Inputs

couchdb_changes
drupal_dblog
elasticsearch
exec
eventlog
file
ganglia
gelf
generator
graphite
github
heartbeat
heroku
irc
imap
jmx
kafka
log4j
lumberjack
meetup
pipe
puppet_facter
relp
rss
rackspace
rabbitmq
redis
snmptrap
stdin
sqlite
s3
sqs
stomp
syslog
tcp
twitter
unix
udp
varnishlog
wmi
websocket
xmpp
zenoss
zeromq

http://www.elastic.co/guide/en/logstash/master/input-plugins.html

Outputs

boundary
circonus
csv
cloudwatch
datadog
datadog_metrics
email
elasticsearch
exec
file
google_bigquery
google_cloud_storage
ganglia
gelf
graphtastic
graphite
hipchat
http
irc
influxdb
juggernaut
jira
kafka
lumberjack
librato
loggly
mongodb
metriccatcher
nagios
null
nagios_nsca
opentsdb
pagerduty
pipe
riemann
redmine
rackspace
rabbitmq
redis
riak
s3
sqs
stomp
statsd
solr_http
sns
syslog
stdout
tcp
udp
websocket
xmpp
zabbix
zeromq

http://www.elastic.co/guide/en/logstash/master/output-plugins.html

Logstash Filters

- Powerful data processing
  - Extract fields from input (grok)
  - Enrich data (geoip, dns)
  - Reformat (csv, split, multiline, json, xml)

Filters

alter	fingerprint	syslog_pri
anonymize	**geoip**	sleep
collate	**grok**	split
csv	i18n	throttle
cidr	**json**	translate
clone	json_encode	uuid
cipher	kv	urldecode
checksum	mutate	useragent
date	metrics	xml
dns	multiline	zeromq
drop	metaevent	
elasticsearch	prune	
extractnumbers	punct	
environment	ruby	
elapsed	range	

http://www.elastic.co/guide/en/logstash/master/filter-plugins.html

Grok — Time to get your RegEx on!

Input data

```
[OracleBIServerComponent] [TRACE:2] [USER-18] [] [ecid: 0000KhjDEIe5UcWFLzESO/
```

```
\[%{DATA:Component}\] \[%{WORD:Severity}(:%{NUMBER:LogLevelNum})?\] \[%{
```

Grok pattern

```
{
  "Component": [
    [
      "OracleBIServerComponent"
    ]
  ],
  "Severity": [
    [
      "TRACE"
    ]
```

Key/Value output

http://grokdebug.herokuapp.com/

Logstash in Action

```
input {file { path => ["nqserver.log" ] }}
```

```
filter { grok { match => [  "message",
"\[%{TIMESTAMP_ISO8601:timestamp}\]
  \[%{DATA:Component}\]
  \[%{WORD:Severity}
(:%{NUMBER:LogLevelNum})?\]
```

```
output {
 elasticsearch {  host => "localhost" }}
```

Logstash -> Elasticsearch

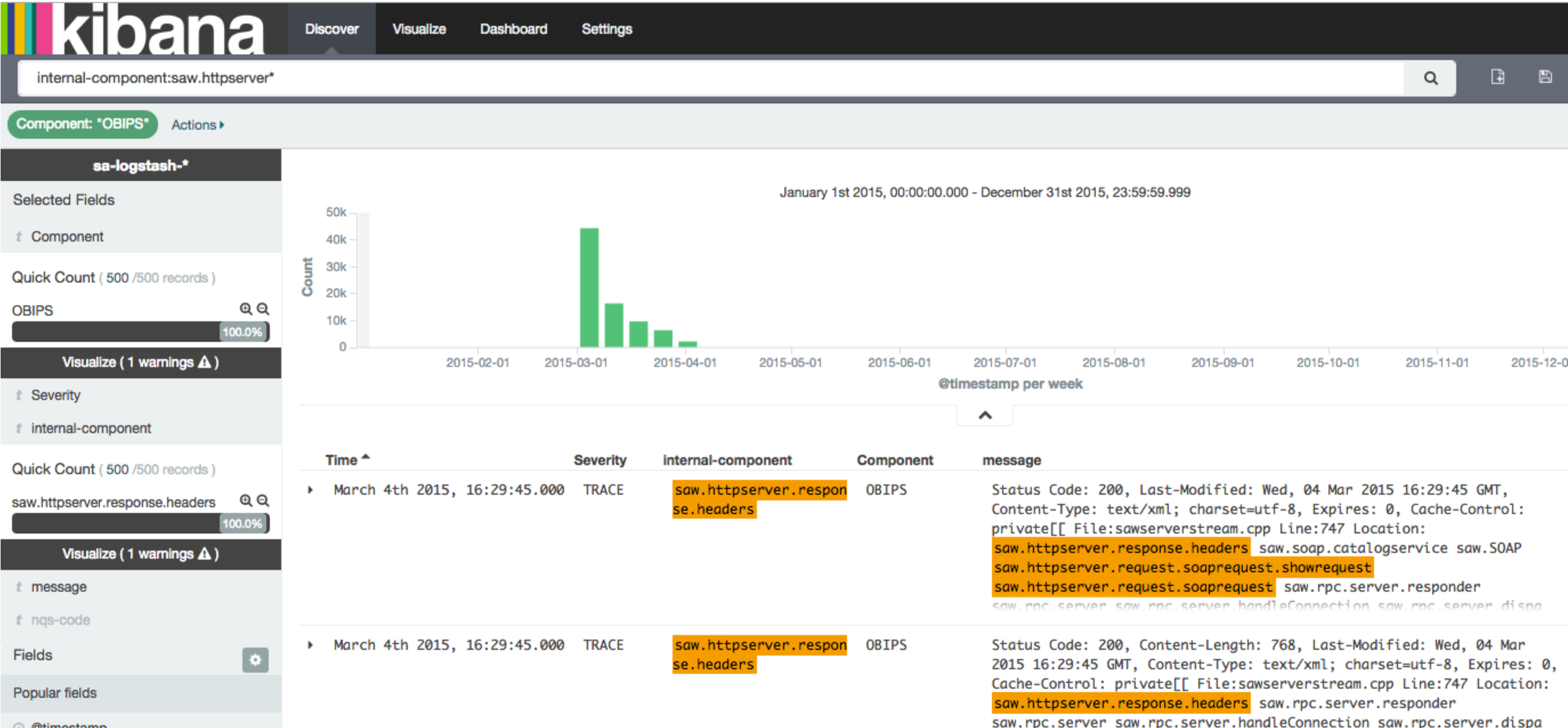
@timestamp	March 28th 2015, 10:16:54.000
@version	1
BMM	01 - Sample App
Component	OracleBIServerComponent
ECID	11d1def534ea1be0:-8528fda:1489e5ec
InternalUser	USER-23
LogLevelNum	3
NQS_SESSIONID	eea0000
REQUESTID	eea0029
RPD	Star
Severity	TRACE
SubjectArea	A - Sample Sales

Systems Diagnostics

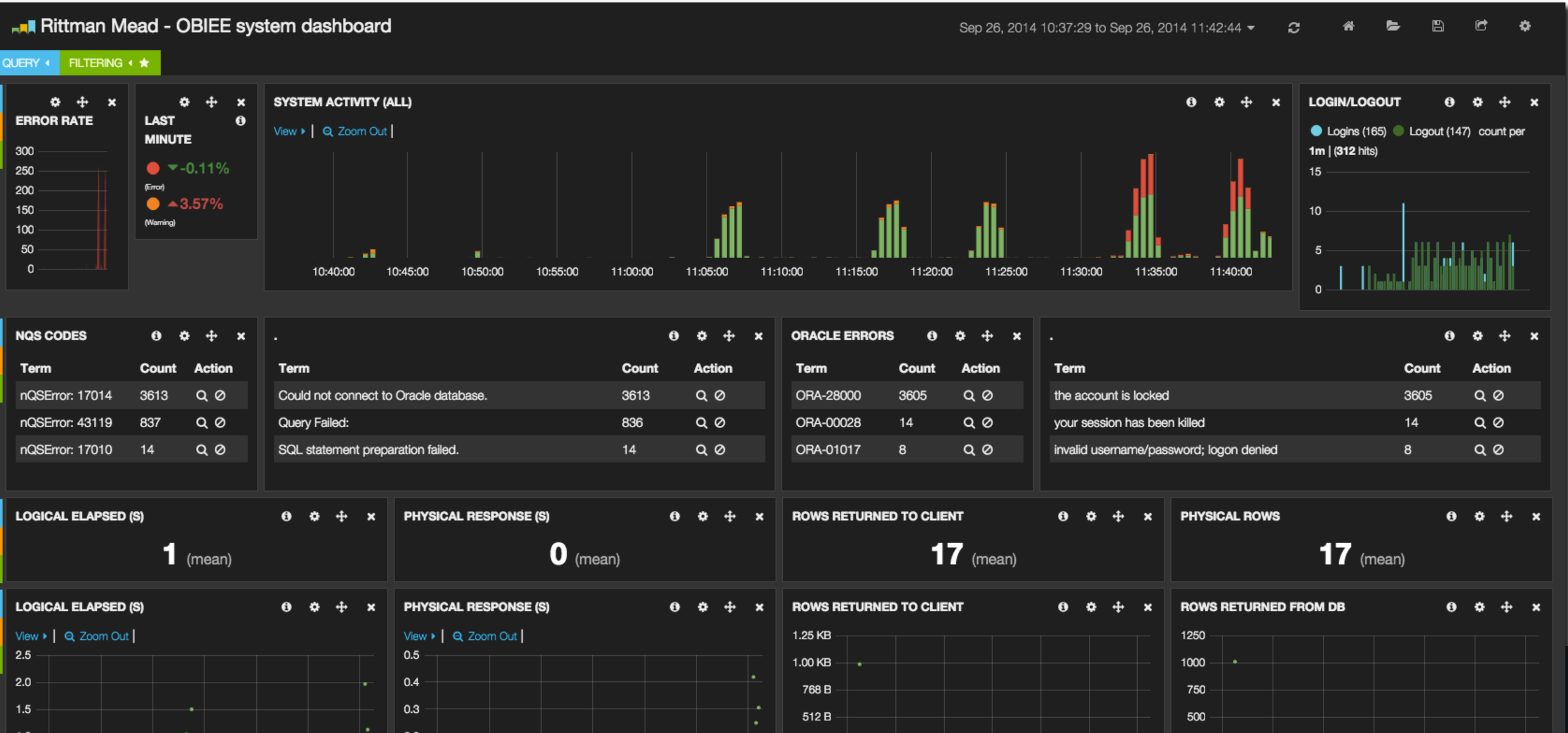
with ELK

[2015-03-25T02:20:31.000-00:00] [OBIPS] [TRACE:1] [] [saw.httpserver.request.servervariables] [ecid: 11d1def534ea1be0:-85
00009f1bdd,0:1] [tid: 3854489344] CONTENT_TYPE=text/xml; charset=utf-8, HTTP_CHUNKED_AWARE=1, HTTP_COOKIE=, IS_SECURE=0,
HOST=demo.us.oracle.com, SCRIPT_NAME=/saw.dll, SERVER_NAME=localhost, SERVER_PORT=7780, SERVER_PROTOCOL=HTTP/1.1, URL=/an
File:rpchttpproviders.cpp
Line:128
Location:
        saw.httpserver.request.servervariables
        saw.rpc.server.responder
        saw.rpc.server
        saw.rpc.server.handleConnection
        saw.rpc.server.dispatch
        saw.threadpool.socketrpcserver
        saw.threads
ecid: 11d1def534ea1be0:-8528fda:1489e5ec29e:-8000-00000000009f1bdd,0:1
ThreadID: 3854489344
]]
[2015-03-25T02:20:31.000-00:00] [OBIPS] [TRACE:1] [] [saw.httpserver.request.soaprequest.showrequest] [ecid: 11d1def534ea
0-00000000009f1bdd,0:1] [tid: 3854489344] Soap Request received.
Type: Headers:
Accept=text/xml, multipart/related, text/html, image/gif, image/jpeg, *; q=.2, */*; q=.2
Connection=Keep-Alive
Content-Length=1185
Content-type=text/xml; charset=utf-8
Host=localhost:7780
Soapaction="#getSessionEnvironment"
User-Agent=Oracle JAX-WS 2.1.5
           Body:
<?xml version='1.0' encoding='UTF-8'?><S:Envelope xmlns:S="http://schemas.xmlsoap.org/soap/envelope/"><S:Header><work:Wor
acle.com/weblogic/soap/workarea/">rO0ABXoAAAJDADBvcmFjbGUuZG1zLmNvbnRleHQuaW50ZXJuYWwud2xzLldMU0NvbnRleHRGYW1pbHkAAAFPAAA
mlhbGl6YWJsZVdvcmtDb250ZXh0AAABSaztAAVzcgAxd2VibG9naWMud29ya2FyZWEuU2VyaWFsaXphYmxlV29ya0NvbnRleHQkQ2Fycmllcv1CAh9z5wpPAw
YWJsZXQAFkxqYXZhL2lvL1NlcmlhbGl6YWJsZTt4cHcEAAAAAXNyAEFvcmFjbGUuZG1zLmNvbnRleHQuaW50ZXJuYWwud2xzLldMU0NvbnRleHRGYW1pbHkkU
AAwABTAARbVdMU0NvbnRleHRGYW1pbHl0ADJMb3JhY2xlL2Rtcy9jb250ZXh0L2ludGVybmFsL3dscy9XTFNDb250ZXh0RmFtaWx5O3hwdykAJzEuMDAwMEtZ
AwMDAwMjt2MHh3AQF4ACZ3ZWJsb2dpYy5kaWFnbm9zdGljcy5EaWFnbm9zdGljQ29udGV4dAAAAX8AAAAyd2VibG9naWMuZGlhZ25vc3RpY3MuY29udGV4dC5
AAiMDAwMEtZVU5oSzJGdzAwMGp6d2tubzFLODViQjAwMDAwMgAAAAIAAAAAAAAA</work:WorkContext></S:Header><S:Body><getSessionEnvironme

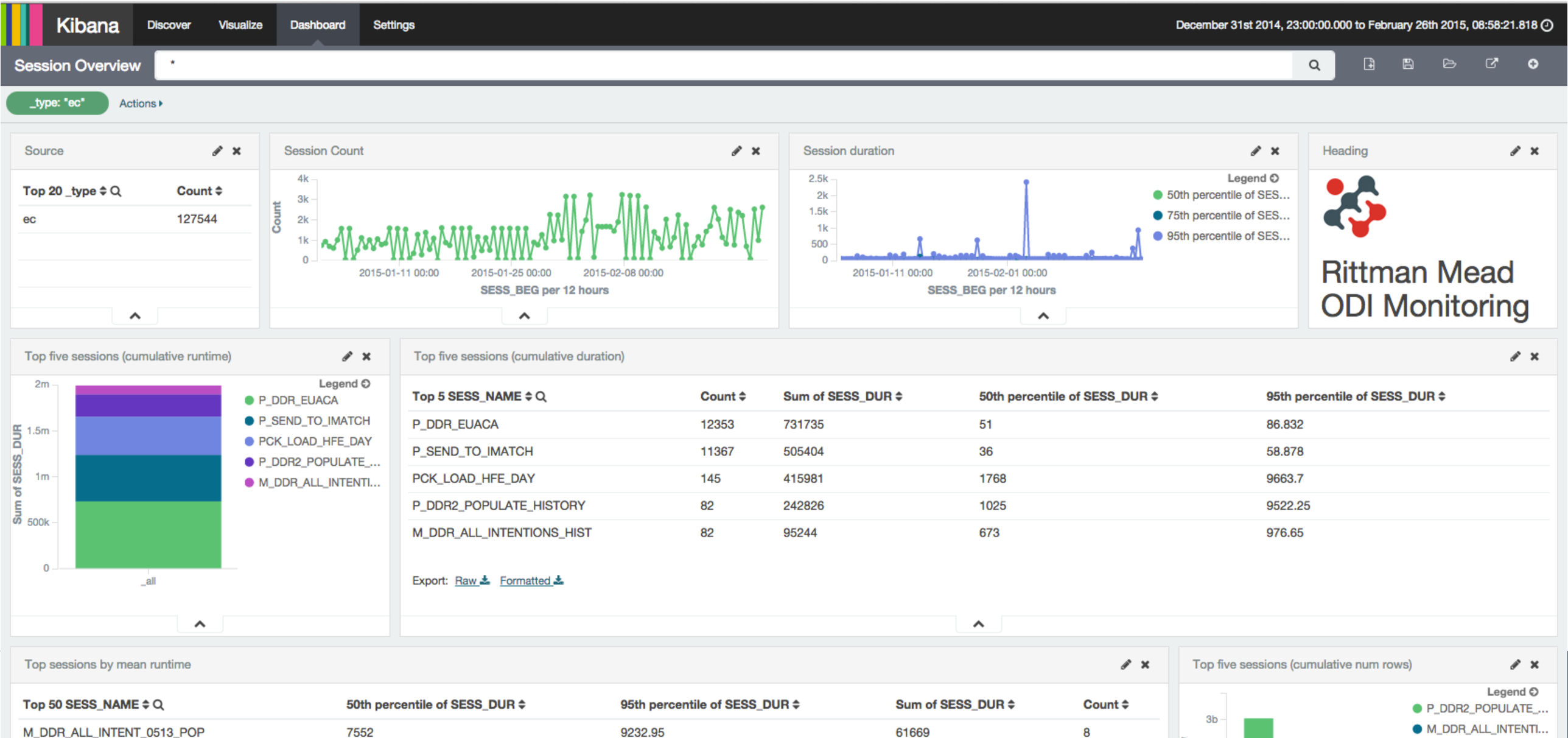
System Diagnostics



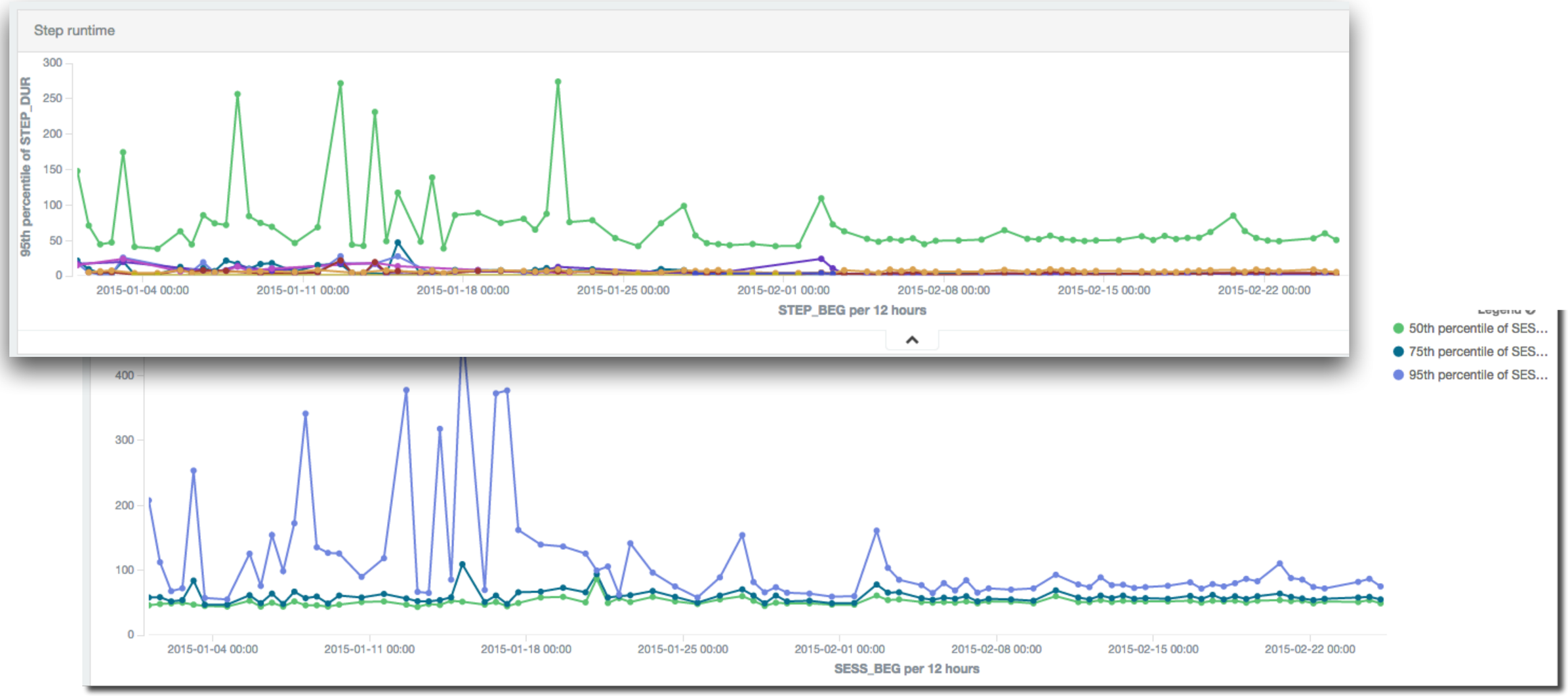
System Monitoring



System Monitoring



Performance Diagnostics



#EOF

Data Discovery

http://ritt.md/go-elk-1

System Diagnostics & Monitoring

http://ritt.md/go-elk-2

http://ritt.md/go-elk-3

Logstash & Kafka

http://ritt.md/kafka-elk

http://ritt.md/kafka-pipelines

email

robin@rittmanmead.com

web

http://ritt.md/rmoff

twitter

@rmoff

irc

rmoff @ #obihackers